



Studying the Mental Effort in Virtual Versus Real Environments

Tiffany Luong, Nicolas Martin, Ferran Argelaguet Sanz, Anatole Lécuyer

► To cite this version:

Tiffany Luong, Nicolas Martin, Ferran Argelaguet Sanz, Anatole Lécuyer. Studying the Mental Effort in Virtual Versus Real Environments. VR 2019 - 26th IEEE Conference on Virtual Reality and 3D User Interfaces, Mar 2019, Osaka, France. pp.809-816, 10.1109/VR.2019.8798029 . hal-02295331

HAL Id: hal-02295331

<https://inria.hal.science/hal-02295331>

Submitted on 24 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Studying the Mental Effort in Virtual Versus Real Environments

Tiffany Luong^{*†}

Nicolas Martin^{*}

Ferran Argelaguet[†]

Anatole Lécuyer[†]

^{*}IRT b<>com, Cesson-Sevigne, France

[†]Univ Rennes, Inria, CNRS, IRISA, Rennes, France

ABSTRACT

Is there an effect of Virtual Reality (VR) Head-Mounted Display (HMD) on the user's mental effort? In this paper, we compare the mental effort in VR versus in real environments. An experiment ($N=27$) was conducted to assess the effect of being immersed in a virtual environment (VE) using an HMD on the user's mental effort while performing a standardized cognitive task (the well-known N -back task, with three levels of difficulty, $N \in \{1, 2, 3\}$). In addition to test the effect of the environment (i.e., virtual versus real), we also explored the impact of performing a dual task (i.e., sitting versus walking) in both environments on mental effort. The mental effort was assessed through self-reports, task performance, behavioural and physiological measures. In a nutshell, the analysis of all measurements revealed no significant effect of being immersed in the VE on the users' mental effort. In contrast, natural walking significantly increased the users' mental effort. Taken together, our results support the fact there is no specific additional mental effort related to the immersion in a VE using a VR HMD.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality

1 INTRODUCTION

The field of Virtual Reality (VR) has recently known an important economical and technological revolution. In particular, VR Head-Mounted Displays (HMDs) are now massively disseminated, and they can be used in a wide spectrum of applications, such as for entertainment, training, or education purposes.

The recent advances in HMD technology enable more and more users to experience impressive feelings of "immersion" and "presence" inside virtual worlds displayed in VR [32]. At the same time, these users often continue to perceive and think that this world they visualize is not real. Such a biased perception, but also the relative cumbersomeness of HMDs, the usual absence of user's virtual body, and other visual artifacts or mismatches with reality could actually require additional information processing from the user, which in turn might increase his/her mental effort.

As such, one question arises regarding the additional workload potentially brought by wearing an HMD: does the fact of being immersed in a virtual environment (VE) while wearing a VR HMD alter the user's cognitive resources? In other words, is the mental effort reported by people immersed in a VE using an HMD the same than in the real world?

In this paper, we aim to assess the potential effect of being immersed in a VE through an HMD on users' mental effort. We conducted a study in which 27 participants did an auditory standardized cognitive test following 3 different levels of difficulty, both in the real world and while wearing an HMD. Since users are rarely static in VR, the effect of a basic secondary task was also tested: our

participants did the cognitive task both in a passive context (while sitting) and in an active context (while walking). Measures of the mental effort were supported by task performance, behavioural, and physiological indicators. Our results showed that being immersed in a VE using a VR HMD does not have an effect on the reported mental effort nor on the cognitive task performance. In contrast, natural walking increased the mental effort.

The remainder of the paper is structured as follows. Section 2 provides an overview of related work regarding the influence of VR on cognitive states. Section 3 presents the experiment we conducted to assess the mental effort in VR versus in reality along with its results. The latter are discussed in Section 4. Finally, Section 5 provides the concluding remarks.

2 RELATED WORK

This section aims at presenting related work on mental effort in VR. It should be noted that while our paper focuses on the mental effort, the latter can be considered as a component of the mental workload [14, 27], which is related to task performance and can sometimes be referred as mental load, mental effort, cognitive load, or cognitive effort in the literature. First, studies dealing with the influence of VR parameters on the mental effort and the task performance will be treated. Then, main methods used to measure mental workload and mental effort in VEs will be depicted.

2.1 Mental effort and task performance in VR

Cognitive states such as the mental effort are of high interest for many fields including VR. They can notably be used as offline metrics to evaluate and improve systems or to study users. Among cognitive theories, one is especially popular: the Multiple Resources Theory of Wickens [40]. This theory proposes that human resources do not have only one information processing, but multiple resources which can be exploited simultaneously or sequentially depending on their type. Related to this theory, the mental workload, which can be defined as "the ratio of demand to allocated resources" [11], has been studied extensively. One of its component, the mental effort, was found to be particularly representative of it [41].

When comparing VR and the real world in terms of mental workload, previous researches have mostly focused on VR training efficiency and on task performance. VR technology has indeed been used extensively to design training applications. In that sense, it was shown that the immersion allowed by VR technologies contributed to improve task performance [2, 4, 23]. Moreover, it was found that skills acquired in VR could be transferred in the real world [9, 29, 31, 39]. In particular, Waller et al. studied the transfer of spatial knowledge from VR to the real world [39]. They found that longer exposure in immersive VEs training could lead to similar performance than training in the real world in a navigation task. Their results also showed that immersive VEs were not efficient to allow subjects a mental representation of the maze map. This was discussed as maybe an effect of the novelty of the HMD and tracker devices on the cognitive effort. Moreover, Chao et al. conducted an experiment to study the effect of nonimmersive VR-system training methods, compared to more traditional methods, in terms of mental workload and training performance [9]. They tested this effect on

^{*}e-mail: firstname.lastname@b-com.com

[†]e-mail: firstname.lastname@inria.fr

tasks related to the maintenance of car pieces, and found VR training method to be more efficient in terms of task performance than the other methods. However, no effect was found on mental workload. It should be noted that they decided not to use full immersive VR systems (such as HMD), as they supposed them to be more likely to cause side-effects such as cybersickness, which may have influenced the user's cognitive effort.

Few works have studied the effect of VR variables and methods on the user's cognitive state. Two studies found an effect of paradigms used in VR on the user's cognitive/mental load. Thereby, it was found that redirected walking influences verbal and spatial working memory performance [6], and that the control of a self-avatar allows better performance in a letter recall task [33]. In addition, the cognitive tasks influenced in both case the user's behaviour. These two studies showed that paradigms used in VR can lead to additional cognitive demands and to changes in behaviour.

2.2 Measures of Mental Workload and Mental Effort

In this paper, the mental effort will be assessed. However, mental effort can be considered as an important index of mental workload [25,27,41] and the methods used to measure both of them can be very similar [21,41]. As there is a larger literature on methods used to measure mental workload than on methods used to measure mental effort, this section will mainly depict methods used to measure mental workload. More details can be found in reviews and surveys on the topic [11,20–22].

O'Donnell classified the methods used to measure mental workload in three groups [24]: subjective, task performance, and physiological measures. Besides these methods, behavioural measures will also be treated since the latter was also proven to be influenced by mental workload [5].

Among subjective measures (here, self-report methods), there are many standardized questionnaires which are commonly used to assess mental workload. Those can be categorized into multidimensional or unidimensional scales. The most popular multidimensional scales are the NASA Task Load Index [14], the Subjective Workload Assessment Technique [27], and the Workload Profile [37], which all consider the mental workload from a multidimensional point of view. Among unidimensional scales, the Rating Scale of Mental Effort (RSME) [41] and the Instantaneous Self Assessment (ISA) [36] are the most used. Due to the decreased number of questions, unidimensional questionnaires are quicker to respond but they can have a smaller diagnosticity as they consider only one dimension [11]. While the ISA and the RSME have been widely validated [8,11], the RSME proved to be one of the most sensitive measure of mental workload in short periods of time [10,38].

Regarding physiological measures, a number of studies have shown correlations between mental workload and physiological signals such as: electroencephalogram (EEG) [5,34], electrooculography (EOG), pupillometry, cardiac activity and electrodermal activity (EDA) signals [3,38]. While EEG and eye-related data have shown good results in the assessment of mental workload [16], those tend to be cumbersome compared to cardiac and EDA sensors [5].

Finally, on the performance and behavioural side, both are dependent on the type of task used to induce the mental workload (i.e. for a binary choice task, a performance measure could be the ratio of correct answers, and a behavioural measure could be the response time [5]).

All these measures can be combined in order to assess the mental effort. Regarding the main advantages of these methods, self-report methods enable true-face validity. On the other hand, performance, physiological and behavioural measures are continuous and do not require a formulated response by the subjects.

In summary, a number of studies have proved that an immersive

VE can improve task performance compared to a less immersive VE. Moreover, works which studied VR training showed that skills acquired in VR could be transferred in the real world. Besides, other studies have emphasized the potential effect of using a full immersive system, such as a HMD, on the user's mental workload or cognitive effort [9,39]. Other works found that paradigms used in VR influenced the user's cognitive load [6,33]. However, the influence of VEs using a VR HMD on the user's subjective mental effort in a pure mental task was not investigated to our knowledge yet.

3 EXPERIMENT

A large amount of studies carried out in VR have considered tasks which required the involvement of motor processes, focusing for instance on the influence of VR training on task performance. There have been way less experiments targeting the user's subjective mental effort in VR using a cognitive task. While task performance is greatly correlated to mental workload, it also greatly depends on the user's motor skills and may not be representative of the user's mental effort if the task is not purely cognitive. Besides, depending on their engagement, some users can show a great performance on a task while involving great cognitive resources, as the opposite is possible. Assessing the user's mental effort seems relevant to study the effect of a VR application to the long term.

More specifically, to the best of the authors knowledge, the effect of being immersed in a VE through wearing an HMD on the user's mental effort has not been investigated yet. A controlled study using a cognitive task and comparing the mental effort reported in both a VE and in the real world would help to answer the question: "Does being immersed in a VE while wearing a VR HMD alter the user's mental effort?". In addition, parting the effect of a simple motor task such as walking on the mental effort during an auditory cognitive task can be interesting since users are rarely static in VR.

3.1 Apparatus

The experiment was conducted in a room of 4.5m x 5m. The room was modelled in 3D using realistic textures, as illustrated in Figure 1. A special attention was given to the design of the virtual environment to make it correspond as much as possible to the real one. The light in the room was controlled using curtains and similar lights layout, intensity and colors. Small objects such as wires were not modelled. Otherwise, physical furniture and markers placed on the floor had its virtual counterparts in the virtual environment, which were mapped correspondingly. The orange rectangle drawn on the floor provided a visual indication to the user regarding the zone where he could walk freely without encountering any obstacle.

Only the trigger button of the Vive controller was used to interact with the support application throughout the experiment. In addition, the pressure exerted on the button was recorded using a 1.8-cm diameter Force Sensitive Resistance (FSR) glued on the trigger button. A wireless module was placed on the Vive controller to allow the streaming of the FSR data on the computer (see Fig. 2). The total weight of the module was around 150g.

In the VR conditions, participants wore an HTC Vive HMD, which provides a resolution of 1200x1080 pixels per eye with a refresh rate of 90Hz. Two eye-tracking devices were used: the Pupil Labs Add-On which was integrated to the HMD, and the SMI eye-tracking glasses. In order to assess the physiological responses of the participants, Blood Volume Pulse (BVP), using PhotoPlethysmography (PPG), and Electrodermal Activity (EDA) were recorded using the Shimmer3 GSR+ [7] in all conditions. The Shimmer module was placed on the forearm of the non-dominant hand (see Fig. 2). The EDA sensors of the Shimmer were placed on the middle phalanx of the non-dominant hand two first fingers, and the BVP sensor, on the earlobe using the Shimmer ear clip. All data was recorded at 120Hz.



Figure 1: (Left) Photo of the room in which the experiment took place. (Right) Screenshot of the virtual environment representing the same room modelled in 3D for the VR conditions.

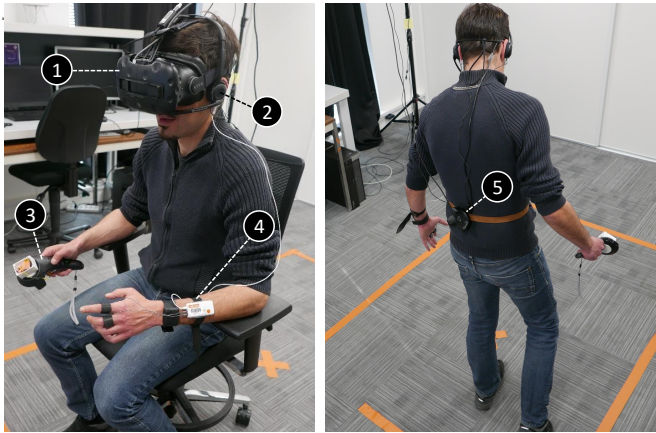


Figure 2: Photo of the experimental set-up in the sitting condition (on the left) and in the walking condition (on the right). (1) HTC Vive HMD (VR) or SMI glasses (Real), (2) Micro-Headset, (3) Vive Controller with a Force Sensitive Resistance sensor on the trigger button, (4) Shimmer3 GSR+, (5) Vive tracker.

Audio instructions were provided using a Sennheiser USB micro-headset. Some visual instructions, such as the announcement of the difficulty, and a reminder of the task principle for each difficulty in the form of an animation were given on the TV screen. The visual presentation was presented in the same way on the modelled TV in the VR conditions than on the real screen in the Real conditions.

As some of the used sensor were not wireless, the wires of the HTC Vive, the SMI glasses and the micro-headset were passed through a ring suspended from the ceiling. The length of the wires was adjusted each time before the walking conditions to match the user height. The cables were maintained in the user's back with a belt so they would not interfere with the user movement. A Vive Tracker was also attached to the belt to track user's movements during the walking conditions (see Fig. 2).

The support application was developed in Unity 3D, and run on a desktop computer equipped with an Intel Xeon CPU E5-1620 v4 processor, one Nvidia GeForce GTX 1080 graphic card, and 16Go Random-Access Memory.

3.2 Participants

29 healthy participants from our research institute volunteered to take part in this study. One subject was excluded from the study due to technical issues, and another one, because of omissions in responses, resulting in a final sample of 27 participants (10 females,

17 males; ages 20-59, $M=32.8$). Two participants wore glasses, six participants wore contact lenses, and one participant reported a slight red-green weakness. One of the participant was left-handed, and all the others, right-handed. Three participants were at least regular users of HMD, and all others had no experience or a slight experience with HMDs in the past. All participants were fluent in French and were naive to the experiment conditions and purpose. They all completed and signed an informed consent form before the start of the experiment. The total time per participant, including pre-questionnaires, instructions, training, experiment, breaks, post-questionnaire, and debriefing, ranged between 55 min and 1 hour 20 minutes. They wore the HMD for about 20 minutes and were allowed to take breaks at any time.

3.3 Experimental Design

As discussed above, exerting cognitive tasks in an immersive VE while wearing an HMD might require more cognitive resources than in the real world. The reported mental effort during natural walking in VR and in the real world might as well differ due to the cumbersomeness of the devices, the differences in perception between the real and the virtual environments, the lack of a virtual body [17] and other VR-related cognitive and perceptive phenomena.

The experiment followed a 3-factor within-subject design. The factors were the difficulty of the auditory cognitive task (3 levels), the environment (VR vs. Real), and the secondary task (Sitting vs. Walking). Each condition was repeated twice, resulting in a $3 \times 2 \times 2 \times 2$ within-subject experimental design. To minimize the learning effect, the order of the environment and task conditions were counterbalanced using a Latin square design, and the difficulty of the cognitive task conditions was randomized in each repetition (see Fig. 3).

3.3.1 Independent variables

In this experiment, three independent variables were considered: Difficulty of the Cognitive Task, Environment, and Secondary Task.

Cognitive task (1- vs. 2- vs. 3-back task)

An auditory N -back letter task was chosen to induce different levels of mental workload [1]. For each sequence of letters, the participants were instructed to press down the trigger button of the Vive controller, if (and only if) the current letter was the same as the N -th previous one and before the next letter presentation. The value of N determined the number of letters the participants had to remember, thus, the difficulty of the task [1]. Three values of N were chosen: 1, 2, and 3 as they are the most common difficulty levels in experiments relying on the N -back working memory paradigm [1]. We chose the 3-back task over the 0-back task since the latter

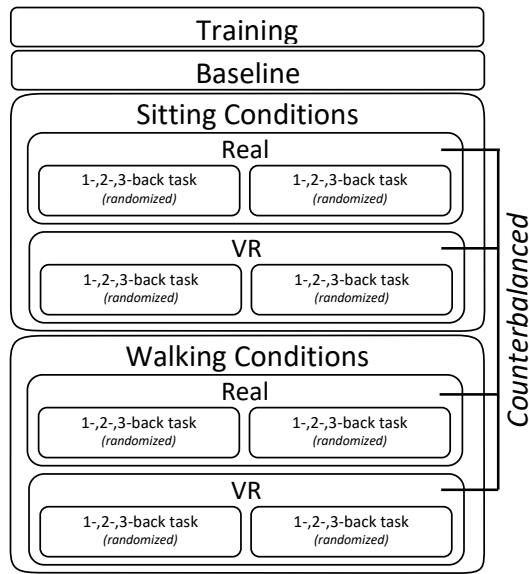


Figure 3: Process of the experiment for one subject (excluding pre, post-questionnaire, instructions, set-up, and calibrations).

mostly is used as a control condition [1], and may not elicit enough difference in mental workload with the 1-back task.

The sequences of letters were made of 3 distinct letters: "A", "B", and "C", which were spoken in French by a male synthetic voice¹. Each sequence of letters was $25 + N$ letters long. A letter could not appear more than $N + 2$ consecutive times as it would have resulted in a fluctuation of the difficulty. Out of the 25 letters which were relevant (the first N were irrelevant as there is no N -th previous one to compare to), 12 letters were the same as the N -th previous one. Consecutive letters were spaced in time by a constant interval of 2.0s (as in [15]), which resulted in a total duration ranged from 52s to 56s for each sequence of letters (depending on the difficulty).

Environment (Virtual Reality vs. Real World)

The participants performed the cognitive tasks in both the Immersive VR (VR), wearing the Vive HMD, and in the real world (Real). A virtual representation of the room in which the experiment took place was given in the VR conditions (see Fig. 1). All visual stimuli were presented on the TV (real or virtual), and the sizes and positions of the virtual furniture and markers matched the real ones. In the sitting conditions, the participants were positioned in front of the TV at a distance of 1.5m.

Secondary Task (Sitting vs. Walking)

In order to see if a basic task had an impact on the reported mental effort in both environments, participants were asked to perform the N -back tasks while sitting and while walking.

In the sitting conditions, participants could blink, look down, fixate a point, but they were asked not to keep their eyes closed all trial long. Aside from this, they were not given any other specific instruction.

In the walking conditions, participants were asked to stay inside the orange rectangle drawn on the floor and to avoid walking on the markers. The participants were asked to start all trials (difficulty) behind the orange line, and to start walking forth in circle around the orange cross at the start of the cognitive task (when they heard the beep) until the end of the task. It was mentioned they could walk at the pace they felt natural and in the direction they wanted for the

first difficulty of the condition (Walking-Real, Walking-VR). After having reported their mental effort, they were asked to alternate the direction in which they were walking between each trial to avoid ordering effects.

3.3.2 Collected Data

The dependent variables considered were: self-reports, which refer to the subjective mental effort reported by the participants, cognitive task performance, behavioural, and physiological measures.

Self-Reports: As this experiment includes many conditions, we decided to orient our choices toward an unidimensional scale: the Rating Scale of Mental Effort (RSME) [41], which is particularly sensitive over short periods of time [10, 38]. The latter evaluates the mental effort invested in a task on a continuous vertical axis from 0 to 150 units (see Fig. 4). Along the vertical axis, some landmarks are labelled with a verbal descriptor of effort, ranging from "absolutely no effort" to "extreme effort" (see Fig. 4).

Performance and Behavioural Measures: Since the primary task was the N -back task, performance indicators of the success of the cognitive task were recorded. We chose to focus on the accuracy value, which corresponds to the number of true responses (true positive and true negative), divided by the total number of responses (true and false values, here, 25). For each trial, the participants should have pressed the trigger button 12 times, and not answered 13 times.

Aside from the performance linked to the cognitive task, we recorded the participants' click behaviour during the trials via the response time. The exerted pressure on the trigger button was also assessed, as it has been shown to be influenced by the arousal (which corresponds to the excitation, mental activation or level of alertness [19]) [30]. In addition, mental workload and mental effort are related to the arousal [22].

Other behavioural measures were recorded during the walking conditions, such as the position and the velocity of the participants.

Physiological Measures: For all conditions, BVP and EDA data were recorded and specific physiological features were extracted. From the BVP signal, the Heart Rate (HR) was computed based on a heartbeat detection. From the EDA signal, the Skin Conductance Level (SCL) was estimated using a low-pass filter (cutoff frequency = 0.05Hz, order = 3).

Since this study intends to compare VR and real conditions, we assumed that the use of two separated eye-tracking devices for each of these conditions would bias the analysis. Thus, eye-tracking data was not considered in this paper.

3.3.3 Experimental procedure

The experiment had a total duration of around 1h10 and was subdivided into the following steps:

Written Consent and Instructions: The informed consent form was given in an email sent to recruit the participants. Those had the possibility to read it prior to the experiment, and to bring it completed and signed at the start of the experiment. Otherwise, they were asked to read it, to complete it and to sign it. They were then instructed with the nature of the experiment, the equipment used, the data recorded (which was anonymized), the VR setup, the cognitive task instructions, and the proceedings of the experiment. Participants were also asked to fill a questionnaire (experience with VR, level of alertness, state of vision, demographic information) to gather information about their background.

Training: Users were then equipped with a Vive Controller and a micro-headset. The RSME was displayed on the TV screen, and

¹Text To Speech. <http://www.fromtexttospeech.com/>. Accessed: 2018-10-25

participants were asked to try to use the Vive controller to evaluate their level of mental effort. They were asked if they well understood the instruction and terms used on the scale. Once participants took control of the interaction, they could perform once every difficulty of the *N*-back task (1-, 2-, 3-back) to make sure they fully understood the instructions of the cognitive task. The training was performed in a Sitting configuration in the Real-World.

Equipment Set-Up and Baseline: At the end of the training, participants were equipped with physiological sensors, aside from the eye tracking devices. They were then asked to remain seated and to relax while the TV screen was black for 1 minute, to record their physiological signals in a relaxed state.

Experiment: The experiment was divided in two parts: the Sitting condition and the Walking condition. Each of these two secondary task conditions (absence or presence of a secondary task) was associated with the environment conditions (Real and VR). The order of these conditions was counterbalanced using a Latin square design. Prior to the Walking group of conditions, the users were instructed with the Walking instructions (see 3.3.1). Each condition lasted around 8 minutes and was preceded by a 2 min calibration of the eye-tracking device (SMI glasses or Pupil Labs Add-on). Next, the participants performed the *N*-back task in two sets of randomized difficulties (6 trials per condition, see Fig. 3). They had the possibility to make a break at anytime and they were proposed to do it before each new condition. The participants were asked to do their best throughout the experiment.

A trial consisted of the following elements (see Fig. 4). It started with a visual instruction on the TV screen, informing the participants which difficulty (1-, 2-, or 3-back) was about to start. After the user pressed the trigger button, an audio announcement of the difficulty was sent to the user, the sequence of letters was initialized, and a visual and animated reminder of the instructions for the current difficulty was displayed. Once the user pressed the trigger button again the task started with a beep followed after 1s by the stream of letters. By the end of the sequence the user was asked to answer the RSME by using the Vive controller. A pointer indicated the direction in which the Vive controller was pointing at on the TV, and the user could move a cursor on the vertical axis by dragging and dropping it. The "+" and "-" buttons could also be used to readjust the value of the reported mental effort. A click on the "Next" button marked the end of the trial, and the beginning of a new trial or the end of the current condition (Sitting-Real, Sitting-VR, Walking-Real, Walking-VR).

Debriefing: At the end of the experiment all sensors were removed from the participants. They were asked to fill a post-questionnaire regarding their subjective impressions. They were finally debriefed and invited to ask all questions they may have had.

3.3.4 Hypotheses

This study aims to evaluate the interaction between cognitive task, basic secondary task, and VR.

Our main hypothesis concerns the effect of the environment on mental effort. We hypothesize that being immersed in a VE through an HMD will impact mental effort [H1]. Moreover, since users are rarely static in VR the effect of basic and natural task such as walking may have an influence on the user's mental effort difference between the virtual and the real environments. This potential effect has to be investigated. It is supposed that doing a cognitive task while walking will significantly increase the reported mental effort compared to doing a cognitive task in a sitting condition [H2]. For both hypothesis it is supposed the results will be supported by the analysis of all the measured indicators of mental effort considered in this study. To summarize, our hypotheses are:

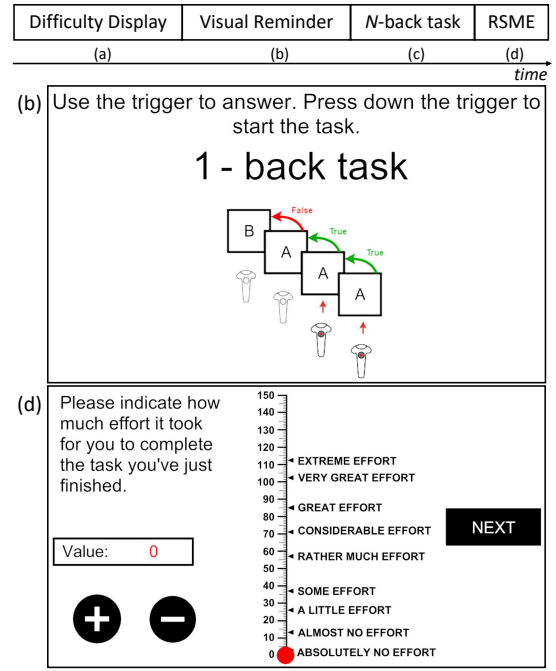


Figure 4: Progress of one trial: (a) Visual display of the task difficulty on the (real or virtual) TV screen ; (b) Auditory announcement of the difficulty, animated GIF reminding the instruction for the difficulty, and initialization of the task (here, for the 1-back task) ; (c) Start of the task with a beep, followed by the sequence of audio letters (N+25 letters) ; (d) Rating Scale of Mental Effort (RSME) [41]. Steps (b) and (c) are triggered by the trigger button, and step (d) appears at the end of the task (step (c)).

- **H1:** Users experience more mental effort in a VE while wearing a VR HMD than in the real world.
- **H2:** Users experience more mental effort while walking than while sitting.

3.4 Results

Generalized linear mixed model (GLMM) analysis was considered for all dependent variables (all parametric). For each variable, the user was considered as a random factor and all the independent variables as within-subject factors. When the equal variances assumption was violated, the degrees of freedom were correction using the Greenhouse-Geisser method. When needed, pairwise post-hoc tests (Bonferroni with adjustment) were performed, only significant differences ($p < 0.05$) are discussed. Order effects were also tested but there are only discussed if significant. The statistical analysis was performed using the R statistical software.

Figure 5 presents the results concerning the self-reported mental effort and the task performance, and Figure 6 presents part of the results on the behavioural and physiological data analysis.

3.4.1 Self-reports

RMSE: The GLMM showed a main effect on Secondary Task $F_{1,26} = 20.81, p < 0.001, \eta_p^2 = 0.44$ and Difficulty $F_{1,76,45,70} = 127.50, p < 0.001, \eta_p^2 = 0.83$, but there was no main effect on Environment $F_{1,26} = 1.09, p = 0.31, \eta_p^2 = 0.04$. Post-hoc tests showed that as the level of difficulty increased, participants perceived the task as more mentally demanding. Moreover, participants perceived the Walking condition to be more mentally demanding (all $p < 0.05$). This results partially support [H2], but do not support [H1].

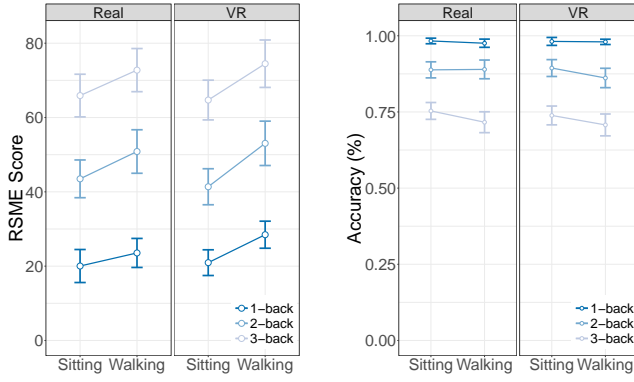


Figure 5: Mean mental effort reported with the RSME and mean accuracy considering all independent variables.

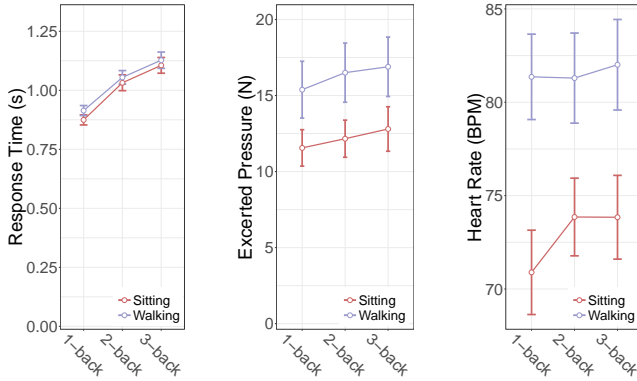


Figure 6: Mean indicators of mental effort (response time, exerted pressure on the button, HR) in function of the secondary task and difficulty. As the environment did not influence significantly these measures, only the effect of the task is displayed.

3.4.2 Task Performance

Accuracy: The GLMM showed a main effect on Secondary Task $F_{1,26} = 7.14$, $p < 0.01$, $\eta_p^2 = 0.22$ and Difficulty $F_{1,80,46,84} = 208.14$, $p < 0.001$, $\eta_p^2 = 0.89$, in contrast there was no effect on Environment $F_{1,26} = 1.26$, $p = 0.27$, $\eta_p^2 = 0.05$. The analysis did not show any significant interaction effect. Similarly as the RSME, post-hoc tests (all $p < 0.05$) showed that as the level of difficulty increased, the accuracy of participants significantly decreased. The analysis of the main effect on Secondary Task showed that the accuracy of participants was higher for the Sitting condition, yet a deeper analysis seems to suggest that this was mainly the case for the highest Difficulty ($N = 3$). This results partially support [H2], but do not support [H1].

3.4.3 Behavioural Measures

Response time: The GLMM analysis showed a main effect for Difficulty $F_{1,75,45,41} = 94.83$, $p < 0.001$, $\eta_p^2 = 0.78$, a marginal effect for Secondary Task $F_{1,26} = 3.93$, $p = 0.058$, $\eta_p^2 = 0.13$ and no effect for Environment $F_{1,26} = 0.19$, $p = 0.66$, $\eta_p^2 = 0.01$. Post-hoc tests showed that the response time significantly increased as the difficulty increases (all $p < 0.05$). This results partially support [H2], but do not support [H1].

Exerted Pressure: Due to some measuring errors, only the data of 21 users could be analyzed. The GLMM analysis showed a main effect for Secondary Task $F_{1,20} = 19.10$, $p < 0.001$, $\eta_p^2 = 0.49$ and Difficulty $F_{1,67,33,38} = 18.72$, $p < 0.001$, $\eta_p^2 = 0.48$, but no effect for Environment $F_{1,20} = 0.48$, $p = 0.5$, $\eta_p^2 = 0.02$. Post-hoc tests showed that the exerted

pressure was higher for the Walking condition, and that for the lowest Difficulty ($N = 1$) participants exerted the lowest force compared to the others ($N = 2$ and $N = 3$). This results partially support [H2], but do not support [H1].

Walking measures: No significant effect was found on the velocity and position data for any of the independent variables.

3.4.4 Physiological Measures

Heart Rate (HR): Due to some recording errors, the analysis was based on the data of 21 participants. The GLMM showed a main effect on Secondary Task $F_{1,21} = 70.03$, $p < 0.0001$, $\eta_p^2 = 0.77$ and Difficulty $F_{1,75,36,67} = 11.67$, $p = 0.0002$, $\eta_p^2 = 0.36$. In contrast there was no effect on Environment $F_{1,21} = 0.08$, $p = 0.77$, $\eta_p^2 = 0.004$. The analysis also revealed a significant interaction effect between Secondary Task and Difficulty $F_{1,56,32,72} = 5.40$, $p = 0.01$, $\eta_p^2 = 0.20$. Post-hoc tests (all $p < 0.05$) showed that the HR of participants significantly increased between the level 1 and 2, but there was no significant difference between the level 2 and 3. The analysis of the main effect on Secondary Task showed that the HR of participants was higher for the Walking condition. The analysis of the interaction effect between Secondary Task and Difficulty showed significant difference on HR between the Difficulty level in Sitting condition (all $p < 0.05$), but not in the Walking condition. This results partially support [H2], but do not support [H1].

Skin Conductance Level (SCL): The GLMM showed a main effect on Secondary Task $F_{1,25} = 5.06$, $p = 0.03$, $\eta_p^2 = 0.17$ ($M_{Sitting} = 4.40$, $SD_{Sitting} = 3.39$; $M_{Walking} = 5.37$, $SD_{Walking} = 3.15$), and no effect of Difficulty $F_{1,98,49,51} = 0.13$, $p = 0.88$, $\eta_p^2 = 0.005$ or Environment $F_{1,25} = 0.28$, $p = 0.60$, $\eta_p^2 = 0.01$. The analysis did not show any significant interaction effect. This results partially support [H2], but do not support [H1].

4 DISCUSSION

4.1 Impact of wearing a HMD in the user's mental effort

Results show that being immersed in a VE while wearing a VR HMD did not differ from the Real World in terms of mental effort in this experiment. Other parameters such as the difficulty of the cognitive task and the presence or not of a basic secondary task were also tested to reinforce the results robustness. While the N -back task showed to be efficient to elicit different levels of mental effort (see Fig. 5), the difference in the responses reported by the participants in the real world and in the corresponding virtual environment was not significant. Moreover, we hypothesized that walking would amplify the potential difference in mental effort reported by the participants. However, no significant effect of the environment was found considering this variable, despite the absence of a virtual body. These results may be explained by the method used to induce the mental effort. In this experiment, the auditory task was well suited as it could be done while walking. However, given the primarily visual nature of VR, it may not have been able to elicit enough difference in mental effort between VR and the Real World. Maybe a primary task relying more on visual stimuli would have led to different results. Also, in our experiment, the cognitive task relied on a verbal working memory paradigm and was compound of letter stimuli. The use of tasks relying on different pools of cognitive resources [40] (i.e. spatial or motor skills) or on different kind of stimuli (i.e. number, shape) might have resulted in different outcomes for the analysis. While the main hypothesis was not conclusive, these results extend the findings made in VR studies regarding cognitive and affective states [12, 13, 18, 28]: virtual experiences can induce similar user responses compared to real ones.

In contrast, the results showed that the walking task had an effect on the mental effort and on the cognitive task accuracy. The effect of the walking condition was especially noticeable for the difficulty ($N = 3$), as seen in Fig. 5. The task was designed to be as natural as possible, considering the technical limitations. It was supposed

not to require additional cognitive resources from the same pools than the ones affected by the mental effort elicited by the N -back task. However, it should be noted that the path constraint may have still influenced the latter. The fact that a simple walking task had an effect on the mental effort means the amount of interaction may need to be quantified in studies assessing similar cognitive states.

Further studies could help understanding the effect of other tasks on the mental effort in VR. For example, it could also be interesting to explore how other parameters might influence the mental effort such as the degree of immersion, the type of VR device (such as CAVEs), the field of view, the frame rate, the exposure time, the screens resolution, the scene lighting or the scene realism.

4.2 Lessons learned

Beside answering our question, this study provides lessons on how to measure mental effort in VR in particular about what worked and what did not. First, the mental effort reported via the RSME was consistent with the subject's accuracy while performing different levels of difficulty of the N -back task (see Fig. 5). This result supports that the RSME can be an appropriate tool in order to measure the level of mental effort between subjects.

Second, the results showed that the level of difficulty had an effect on the HR, the response time, and on the pressure exerted on the trigger button. Since the mental effort was found to be influenced by the difficulty of the cognitive task (see Section 3.4.1), these findings are consistent with the literature [11, 30]. Nevertheless, depending on the context, not all of the indicators were efficient to measure the mental effort. On one hand, the response time was able to discriminate between each three levels of difficulty in the presence and in the absence of the secondary task. On the other hand in the sitting condition both HR and exerted pressure measures were able to differentiate the 1-back task from the 2-back task, but not the 2-back task from the 3-back task (see Fig. 6). This can be explained by the between-individual variability. Most people consider the 1-back task as an easy task, and the 3-back task as a difficult task. However, the perceived difficulty tend to vary concerning the 2-back task, which may cause more subtle changes in the physiological and behavioural measures. In the walking condition, the exerted pressure could still discriminate between the 1-back task and the 2-back task, but the HR was unable to discriminate between any of the difficulties. This outcome is consistent, since movements tend to add noises in the physiological signals and to decrease the distinction between signal features [35].

Finally, while users' body behaviour and gesture showed to be influenced by the cognitive load in past studies, the analysis of the walking behaviour was not successful to show an effect in this experiment. This might be explained by the fact the space in which the subjects could walk in VR and in the real world was restrained. A bigger space and no path constraint may have lead to different outcomes. Also, the analysis of the SCL was not able to discriminate between the different independent variables in this experiment, aside from the task. An explanation could be that the cognitive task required the subject to be focused, and was not specially arousing.

In summary, in our experiment, among behavioural and physiological measures, the response time was the best indicator of mental effort, followed by the exerted pressure, then, by the HR. Cautions should however be taken when using physiological sensors in interactive environments, as their viability tend to greatly decrease with the involvement of movements [35]. Also, while the pressure exerted on a button was shown to be at least as efficient to differentiate the mental effort as the HR in this experiment, it was rarely used in the literature. We therefore encourage the use of FSR sensors in future studies involving the use of interactive buttons to support measures of the cognitive states, as they are cheap, not cumbersome, and easy to put. Finally, machine learning approaches

could be applied in order to assess the mental effort in VR in real-time, based on physiological and behavioural responses [26].

5 CONCLUSION

This paper is the first to specifically study the influence of being immersed in a VE using a VR HMD on the user's mental effort. We conducted an experiment in which participants had to achieve a standardized cognitive task with various levels of difficulty in two different environments (real vs. virtual), and in presence or not of a secondary task (sitting vs. walking). The results showed that wearing an HMD did not significantly impact the user's mental effort and his/her task performance. In contrast, walking significantly increased mental effort and decreased task performance. These findings are grounded on various kinds of converging measurement: self-report, task accuracy, mean heart-rate, response time, and pressure exerted on the answering button.

Taken together, our results support the view that being immersed in a VE while wearing an HMD does not significantly influence the provided mental effort. As such, these results promote the use of VR HMDs in simulations and experimental studies involving measurements of human's cognitive states. Moreover, secondary interactions such as walking should be carefully controlled, as they can potentially influence and add noise in mental effort measurements in VR.

ACKNOWLEDGMENTS

Our study was carried out within b<>com, an institute of research and technology dedicated to digital technologies. It received support from the Future Investments program of the French National Research Agency (grant no. ANR-07-A0-AIRT). We would like to thank Pierre Le Gargasson and Jean-Marc Diverrez for their help with the sensors.

REFERENCES

- [1] J. Au, E. Sheehan, N. Tsai, G. J. Duncan, M. Buschkuehl, and S. M. Jaeggi. Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic bulletin & review*, 22(2):366–377, 2015.
- [2] W. Barfield, C. Hendrix, and K.-E. Bystrom. Effects of stereopsis and head tracking on performance using desktop virtual environment displays. *Presence: Teleoperators & Virtual Environments*, 8(2):237–240, 1999.
- [3] W. Boucsein. Electrodermal indices of emotion and stress, chapter 3. *Electrodermal Activity*, pp. 369–391, 1992.
- [4] D. A. Bowman and R. P. McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7), 2007.
- [5] A.-M. Brouwer, M. A. Hogervorst, J. B. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering*, 9(4):045008, 2012.
- [6] G. Bruder, P. Lubas, and F. Steinicke. Cognitive resource demands of redirected walking. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):539–544, 2015.
- [7] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca. Shimmer™—a wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal*, 10(9):1527–1534, 2010.
- [8] B. Cain. A review of the mental workload literature. Technical report, Defence Research And Development Toronto (Canada), 2007.
- [9] C.-J. Chao, S.-Y. Wu, Y.-J. Yau, W.-Y. Feng, and F.-Y. Tseng. Effects of three-dimensional virtual reality and traditional training methods on mental workload and training performance. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 27(4):187–196, 2017.
- [10] F. P. Da Silva. Mental workload, task demand and driving performance: what relation? *Procedia-Social and Behavioral Sciences*, 162:310–319, 2014.

- [11] D. De Waard. *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center Netherlands, 1996.
- [12] A. Felnhofer, O. D. Kothgassner, T. Hetterle, L. Beutl, H. Hlavacs, and I. Kryspin-Exner. Afraid to be there? Evaluating the relation between presence, self-reported anxiety, and heart rate in a virtual public speaking task. *Cyberpsychology, Behavior, and Social Networking*, 17(5):310–316, 2014.
- [13] A. Felnhofer, O. D. Kothgassner, M. Schmidt, A.-K. Heinzle, L. Beutl, H. Hlavacs, and I. Kryspin-Exner. Is virtual reality emotionally arousing? Investigating five emotion inducing virtual park scenarios. *International journal of human-computer studies*, 82:48–56, 2015.
- [14] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in Psychology*, vol. 52, pp. 139–183. Elsevier, 1988.
- [15] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz. Mental workload during n-back task—quantified in the prefrontal cortex using fmirs. *Frontiers in Human Neuroscience*, 7:935, 2014.
- [16] M. A. Hogervorst, A.-M. Brouwer, and J. B. Van Erp. Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience*, 8:322, 2014.
- [17] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.
- [18] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks Jr. Physiological measures of presence in stressful virtual environments. *ACM Transactions On Graphics (TOG)*, 21(3):645–652, 2002.
- [19] A. Mehrabian and J. A. Russell. *An approach to environmental psychology*. the MIT Press, 1974.
- [20] S. Miller. Workload measures. *National Advanced Driving Simulator*. Iowa City, United States, 2001.
- [21] N. Moray. *Mental workload: Its theory and measurement*, vol. 8. Springer Science & Business Media, 2013.
- [22] G. Mulder. The concept and measurement of mental effort. In *Energetics and human information processing*, pp. 175–198. Springer, 1986.
- [23] M. Narayan, L. Waugh, X. Zhang, P. Bafna, and D. Bowman. Quantifying the benefits of immersion for collaboration in virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 78–81. ACM, 2005.
- [24] R. D. O'Donnell. *Workload assessment methodology*. Wiley, 1986.
- [25] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, 38(1):63–71, 2003.
- [26] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, 23(10):1175–1191, 2001.
- [27] G. B. Reid and T. E. Nygren. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Advances in Psychology*, vol. 52, pp. 185–218. Elsevier, 1988.
- [28] G. Riva, F. Mantovani, C. S. Capideville, A. Preziosa, F. Morganti, D. Villani, A. Gaggioli, C. Botella, and M. Alcañiz. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, 10(1):45–56, 2007.
- [29] F. Rose, E. A. Attree, B. Brooks, D. Parslow, and P. Penn. Training in virtual environments: transfer to real world tasks and equivalence to real task training. *Ergonomics*, 43(4):494–511, 2000.
- [30] K. Schaaff, R. Degen, N. Adler, and M. T. Adam. Measuring affect using a standard mouse device. *Biomedical Engineering/Biomedizinische Technik*, 57(SI-1 Track-N):761–764, 2012.
- [31] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O'Brien, V. K. Bansal, D. K. Andersen, and R. M. Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery*, 236(4):458, 2002.
- [32] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.
- [33] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 67–76. IEEE, 2016.
- [34] M. B. Sterman, C. A. Mann, D. A. Kaiser, and B. Y. Suyenobu. Multi-band topographic EEG analysis of a simulated visuomotor aviation task. *International journal of psychophysiology*, 16(1):49–56, 1994.
- [35] K. T. Sweeney, T. E. Ward, and S. F. McLoone. Artifact removal in physiological signals—practices and possibilities. *IEEE transactions on information technology in biomedicine*, 16(3):488–500, 2012.
- [36] A. J. Tattersall and P. S. Foord. An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, 39(5):740–748, 1996.
- [37] P. S. Tsang and V. L. Velazquez. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3):358–381, 1996.
- [38] W. B. Verwey and H. A. Veltman. Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of experimental psychology: Applied*, 2(3):270, 1996.
- [39] D. Waller, E. Hunt, and D. Knapp. The transfer of spatial knowledge in virtual environment training. *Presence*, 7(2):129–143, 1998.
- [40] C. D. Wickens. Multiple resources and mental workload. *Human Factors*, 50(3):449–455, 2008.
- [41] F. R. H. Zijlstra. *Efficiency in work behaviour: A design approach for modern tools*. PhD thesis, 1993.